

PCT/JP 03/03450

日 本 国 特 許 庁
JAPAN PATENT OFFICE

#2
20.03.03

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office

出 願 年 月 日
Date of Application:

2002年 4月12日

出 願 番 号
Application Number:

特願2002-110259

[ST.10/C]:

[JP 2002-110259]

出 願 人
Applicant(s):

三菱電機株式会社

REC'D 16 MAY 2003

WIPO

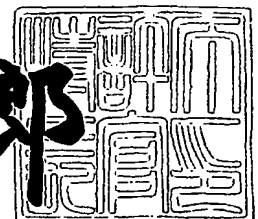
PCT

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH
RULE 17.1(a) OR (b)

2003年 5月 2日

特 許 庁 長 官
Commissioner,
Japan Patent Office

太田 信一郎



出証番号 出証特2003-3031611

【書類名】 特許願

【整理番号】 539982JP01

【提出日】 平成14年 4月12日

【あて先】 特許庁長官 殿

【国際特許分類】 H04N 5/76

【発明者】

【住所又は居所】 東京都千代田区丸の内二丁目2番3号 三菱電機株式会社
社内

【氏名】 守屋 芳美

【発明者】

【住所又は居所】 東京都千代田区丸の内二丁目2番3号 三菱電機株式会社
社内

【氏名】 西川 博文

【発明者】

【住所又は居所】 東京都千代田区丸の内二丁目2番3号 三菱電機株式会社
社内

【氏名】 関口 俊一

【発明者】

【住所又は居所】 東京都千代田区丸の内二丁目2番3号 三菱電機株式会社
社内

【氏名】 山田 悦久

【発明者】

【住所又は居所】 東京都千代田区丸の内二丁目2番3号 三菱電機株式会社
社内

【氏名】 乙井 研二

【発明者】

【住所又は居所】 東京都千代田区丸の内二丁目2番3号 三菱電機株式会社
社内

【氏名】 黒田 慎一

【発明者】

【住所又は居所】 東京都千代田区丸の内二丁目2番3号 三菱電機株式会社
社内

【氏名】 小川 文伸

【発明者】

【住所又は居所】 東京都千代田区丸の内二丁目2番3号 三菱電機株式会社
社内

【氏名】 浅井 光太郎

【特許出願人】

【識別番号】 000006013

【氏名又は名称】 三菱電機株式会社

【代理人】

【識別番号】 100102439

【弁理士】

【氏名又は名称】 宮田 金雄

【選任した代理人】

【識別番号】 100092462

【弁理士】

【氏名又は名称】 高瀬 彌平

【手数料の表示】

【予納台帳番号】 011394

【納付金額】 21,000円

【その他】

国等の委託研究の成果に係る特許出願（平成13年度、
通信・放送機構、情報家電コンテンツ表現システムの研
究開発委託研究、産業活力再生特別措置法第30条の適
用を受けるもの）

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】	要約書 1
【ブルーの要否】	要

【書類名】 明細書

【発明の名称】 メタデータ編集装置およびその方法、メタデータ再生装置およびその方法、メタデータ配信装置およびその方法

【特許請求の範囲】

【請求項1】 動画像と音声のいずれか一方、または双方を含むマルチメディアコンテンツを1つ以上のシーンに分割し、分割したシーン毎にメタデータを生成するメタデータ編集装置であって、

上記1つ以上のシーンから構成されるマルチメディアコンテンツの階層的な構造を記述するシーン構造情報メタデータを生成するシーン構造記述部、

を有することを特徴とするメタデータ編集装置。

【請求項2】 請求項1記載のメタデータ編集装置において、さらに、上記1つ以上に分割された各シーンの中から各シーンがもつ特徴量を抽出し、特徴記述メタデータを生成する特徴抽出部を有することを特徴とするメタデータ編集装置。

【請求項3】 請求項1または請求項2記載のメタデータ編集装置において、さらに、

各シーンの1フレーム以上の代表フレームを画像データとして抽出し、抽出した代表フレーム画像に関するメタデータを生成する代表フレーム生成部を有することを特徴とするメタデータ編集装置。

【請求項4】 請求項1～請求項3いずれかの請求項に記載のメタデータ編集装置において、さらに、

各シーンに対してテキスト情報を付与し、テキスト情報メタデータを生成するテキスト情報付与部を有することを特徴とするメタデータ編集装置。

【請求項5】 請求項1～請求項4いずれかの請求項に記載のメタデータ編集装置において、さらに、生成された各メタデータをMPEG-7に規定されたフォーマットに従い記述するメタデータ記述部を有することを特徴とするメタデータ編集装置。

【請求項6】 請求項1～請求項5いずれかの請求項に記載のメタデータ編集装置において、さらに、シーンチェンジ個所を検出するシーンチェンジ検出部

を有することを特徴とするメタデータ編集装置。

【請求項7】 動画像と音声のいずれか一方、または双方を含むマルチメディアコンテンツを1つ以上のシーンに分割し、分割したシーン毎にメタデータを生成する際、

1つ以上のシーンから構成されるマルチメディアコンテンツの階層的な構造を記述するシーン構造情報メタデータを生成することを特徴とするコンテンツ編集方法。

【請求項8】 動画像と音声のいずれか一方、または双方を含むマルチメディアコンテンツの構造や特徴が規定のフォーマットに従って記述されたメタデータを入力して解析するメタデータ解析手段と、

上記メタデータ解析部の解析結果を入力して、コンテンツの階層的なシーン構造を表示する構造表示部と、

を有することを特徴とするメタデータ再生装置。

【請求項9】 請求項8記載のメタデータ再生装置において、さらに、

上記メタデータ解析部の解析結果によりメタデータから取得されるシーンまたはシーングループの特徴量やテキスト情報に基づいて、シーンまたはシーングループを検索する検索手段を有することを特徴とするメタデータ再生装置。

【請求項10】 請求項8または請求項9記載のメタデータ再生装置において、さらに、

各シーンの重要度に基づいてシーン構造の要約を作成する要約作成部と、

要約作成部によって作成されたシーン構造の要約を表示する要約構造表示部とを有することを特徴とするメタデータ再生装置。

【請求項11】 請求項8～請求項10いずれかの請求項に記載のメタデータ再生装置において、さらに、

表示されているシーン構造の中から選択されたシーンにより構成されるマルチメディアコンテンツを再生するコンテンツ再生部を有することを特徴とするメタデータ再生装置。

【請求項12】 動画像と音声のいずれか一方、または双方を含むマルチメディアコンテンツの構造や特徴が規定のフォーマットに従って記述されたメタデ

ータを入力して解析し、その解析結果を入力して、コンテンツの階層的なシーン構造を表示することを特徴とするメタデータ再生方法。

【請求項 1 3】 動画像と音声のいずれか一方、または双方を含むマルチメディアコンテンツの構造や特徴が規定のフォーマットに従って記述されたメタデータを入力して解析するメタデータ解析手段と、

上記メタデータ解析手段によって解析された情報を入力して、コンテンツを配信すべき各端末の能力に基づきメタデータを再生成するメタデータ再生成部と、

上記メタデータ再生成部によって再生成されたメタデータを各端末に配信するメタデータ配信部と、

を有することを特徴とするメタデータ配信装置。

【請求項 1 4】 請求項 1 2 記載のメタデータ配信サーバにおいて、さらに、メタデータに対応するコンテンツを配信するコンテンツ配信部を有することを特徴とするメタデータ配信装置。

【請求項 1 5】 請求項 1 3 記載のメタデータ配信サーバにおいて、コンテンツ配信サーバは、コンテンツを配信するときに、各端末からの要求条件に応じて、コンテンツのフォーマットを変換して配信することを特徴とするメタデータ配信装置。

【請求項 1 6】 請求項 7 に記載のコンテンツ編集方法により生成されたメタデータを入力して解析し、解析した情報からコンテンツを配信すべき各端末の能力に基づきメタデータを再生成して、再生成したメタデータを各端末に配信することを特徴とするメタデータ配信方法。

【発明の詳細な説明】

【0 0 0 1】

【発明の属する技術分野】

本発明は、動画像や音声を含むマルチメディアコンテンツを複数のシーンに分割し、分割したシーン毎にメタデータを生成するメタデータ編集装置およびその方法と、そのメタデータを用いてコンテンツを再生するメタデータ再生装置およびその方法と、そのために必要なメタデータを配信するメタデータ配信サーバおよびその方法に関する。

【0002】

【従来の技術】

従来技術として、例えば、特開平2001-028722号公報に開示されている動画像管理装置がある。この動画像管理装置では、ビデオデータを複数のシーンに分割し、各シーンに対して、そのシーンの再生に必要な区間情報と、シーン番号と、シーンを代表する画像との集まりであるメタデータを作成し、各インデックスに検索目的に応じたタイトルを付与するインデックス作成手段を有する。ユーザはタイトルを用いてメタデータを検索し、インデックスに含まれるシーン番号の順番と区間情報により、ビデオデータの中の必要なシーンだけをつなぎ合わせて再生することができる。

【0003】

【発明が解決しようとする課題】

しかし、上述の従来技術では、シーンの再生に必要な区間情報と、シーン番号と、シーンを代表する画像とによりメタデータを作成するに過ぎないため、ビデオデータがもっている階層性など、ビデオデータの構造まで管理することができない、という問題があった。

【0004】

また、登録した画像を検索する際には、インデックスに付与されているタイトルを用いて行うため、適切な検索結果を得るためには、適切なタイトルを入力しなければならない、という問題もあった。

【0005】

本発明は、このような問題を解決するためになされたものであり、シーンの区間情報やタイトルのほかに、ビデオデータ等のコンテンツのもつ構造等のインデックス情報であるメタデータを生成することができるメタデータ編集装置およびその方法を提供することを目的とする。

【0006】

また、メタデータ編集装置およびその方法により生成されたメタデータを用いて、ユーザが見たいシーンだけを集めて再生したり、メタデータに記述されている特徴量等を用いて、見たいシーンを検索することができるメタデータ再生装置

およびその方法を、さらにはそのために必要なメタデータを配信するメタデータ配信サーバおよびその方法を提供することを目的とする。

【0007】

【課題を解決するための手段】

上記目的を達成するため、本発明では、動画像と音声のいずれか一方、または双方を含むマルチメディアコンテンツを1つ以上のシーンに分割し、分割したシーン毎にメタデータを生成するメタデータ編集装置であって、上記1つ以上のシーンから構成されるマルチメディアコンテンツの階層的な構造を記述するシーン構造情報メタデータを生成するシーン構造記述部、を有することにより、マルチメディアコンテンツの構造を記述するインデックス情報であるメタデータを生成するメタデータ編集装置であることを特徴とする。

【0008】

特に、さらに、上記1つ以上に分割された各シーンの中から各シーンがもつ特徴量を抽出し、特徴記述メタデータを生成する特徴抽出部を有することにより、さらにマルチメディアコンテンツの特徴を記述するインデックス情報であるメタデータを生成するメタデータ編集装置であることを特徴とする。

【0009】

また、さらに、各シーンの1フレーム以上の代表フレームを画像データとして抽出し、抽出した代表フレーム画像に関するメタデータを生成する代表フレーム生成部を有することにより、コンテンツの内容把握を目的としたインデックス情報であるメタデータを生成するメタデータ編集装置であることを特徴とする。

【0010】

また、さらに、各シーンに対してテキスト情報を付与し、テキスト情報メタデータを生成するテキスト情報付与部を有することにより、検索に有用なインデックス情報であるメタデータを生成するメタデータ編集装置であることを特徴とする。

【0011】

また、さらに、生成された各メタデータをMPEG-7に規定されたフォーマットに従い記述するメタデータ記述部を有するメタデータ編集装置であることを特徴と

する。

【0012】

また、さらに、シーンチェンジ個所を検出するシーンチェンジ検出部を有することにより、人手によらず自動的にシーンチェンジ個所を検出するメタデータ編集装置であることを特徴とする。

【0013】

また、動画像と音声のいずれか一方、または双方を含むマルチメディアコンテンツを1つ以上のシーンに分割し、分割したシーン毎にメタデータを生成する際、1つ以上のシーンから構成されるマルチメディアコンテンツの階層的な構造を記述するシーン構造情報メタデータを生成するコンテンツ編集方法であることを特徴とする。

【0014】

また、動画像と音声のいずれか一方、または双方を含むマルチメディアコンテンツの構造や特徴が規定のフォーマットに従って記述されたメタデータを入力して解析するメタデータ解析手段と、上記メタデータ解析部の解析結果を入力して、コンテンツの階層的なシーン構造を表示する構造表示部と、を有するメタデータ再生装置であることを特徴とする。

【0015】

また、さらに、上記メタデータ解析部の解析結果によりメタデータから取得されるシーンまたはシーングループの特徴量やテキスト情報に基づいて、シーンまたはシーングループを検索する検索手段を有するメタデータ再生装置であることを特徴とする。

【0016】

また、さらに、各シーンの重要度に基づいてシーン構造の要約を作成する要約作成部と、要約作成部によって作成されたシーン構造の要約を表示する要約構造表示部とを有するメタデータ再生装置であることを特徴とする。

【0017】

また、表示されているシーン構造の中から選択されたシーンにより構成されるマルチメディアコンテンツを再生するコンテンツ再生部を有するメタデータ再生

装置であることを特徴とする。

【0018】

また、動画像と音声のいずれか一方、または双方を含むマルチメディアコンテンツの構造や特徴が規定のフォーマットに従って記述されたメタデータを入力して解析し、その解析結果を入力して、コンテンツの階層的なシーン構造を表示するメタデータ再生方法であることを特徴とする。

【0019】

また、動画像と音声のいずれか一方、または双方を含むマルチメディアコンテンツの構造や特徴が規定のフォーマットに従って記述されたメタデータを入力して解析するメタデータ解析手段と、上記メタデータ解析手段によって解析された情報を入力して、コンテンツを配信すべき各端末の能力に基づきメタデータを再生成するメタデータ再生成部と、上記メタデータ再生成部によって再生成されたメタデータを各端末に配信するメタデータ配信部と、を有するメタデータ配信装置であることを特徴とする。

【0020】

また、さらに、メタデータに対応するコンテンツを配信するコンテンツ配信部を有するメタデータ配信装置であることを特徴とする。

【0021】

また、コンテンツ配信サーバは、コンテンツを配信するときに、各端末からの要求条件に応じて、コンテンツのフォーマットを変換して配信することを特徴とする。

【0022】

また、上述のコンテンツ編集方法により生成されたメタデータを入力して解析し、解析した情報からコンテンツを配信すべき各端末の能力に基づきメタデータを再生成して、再生成したメタデータを各端末に配信するメタデータ配信方法であることを特徴とする。

【0023】

【発明の実施の形態】

実施の形態 1.

本実施の形態1では、動画像や音声を含むマルチメディアコンテンツを複数のシーンに分割し、シーンの階層的な構造記述や、各シーンの特徴量を含む本発明のインデックス情報としてメタデータを作成するメタデータ編集装置について説明する。

【0024】

図1は、本実施の形態1におけるメタデータ編集装置の構成を示す。

図1において、2は編集対象のコンテンツを再生・表示するコンテンツ再生・表示部、3はコンテンツを複数のシーンに分割するシーン分割部、4はシーンの代表フレームをサムネイル画像として抽出するサムネイル画像生成部、5はシーン分割部3により分割されたシーンのグループ化、シーンの結合、シーンの削除、シーンの関係情報の生成などによりシーンを階層的に編集するシーン編集部、6は各シーンに対して各種のテキスト情報を付加するテキスト情報付与部、7はシーンの特徴を抽出する特徴抽出部、8はコンテンツ再生・表示部2、シーン分割部3、サムネイル画像生成部4、シーン記述編集部5およびテキスト情報付与部6等に対してユーザからの指示情報を入力するユーザ入力部、9はシーン分割部3、サムネイル画像生成部4、シーン編集部5、テキスト情報付与部6および特徴抽出部7より出力される各種メタデータ情報を統合し、規定のフォーマットに従って記述したメタデータを生成するメタデータ記述部、である。また、10は映像データや音声データ等からなるマルチメディアコンテンツ、11はユーザ入力情報、12はシーンの区間情報メタデータ、13はシーンのサムネイル画像情報メタデータ、14はシーン構造情報メタデータ、15はテキスト情報メタデータ、16は特徴記述メタデータ、17はマルチメディアコンテンツの内容及び構造を記述したメタデータである。

【0025】

次に、本実施の形態1におけるメタデータ編集装置の動作を、図2に示す構成のニュース映像を編集する場合を例に説明する。

【0026】

まず、コンテンツ蓄積部（図示せず）に蓄積されている映像コンテンツ等のマルチメディアコンテンツ10がネットワーク等を介し本装置に入力すると、編集

用にコンテンツ再生・表示部2にて再生・表示される。

【0027】

メタデータ編集装置のユーザがその再生映像を見ながら、シーンとして切り出す位置、すなわちシーンの開始位置と、終了位置とをユーザ入力部8より入力すると、シーン分割部3は、ユーザから入力されたシーンの開始位置と終了位置とを示すシーンの区間情報メタデータ12を生成する。

【0028】

図3に、本実施の形態1のシーン分割部3が図2に示すニュース映像から生成したシーンの区間情報メタデータ12の一例を示す。図3に示すように、シーン分割部3によって、図2に示すニュース映像コンテンツから切り出された“ニュースダイジェスト”や“国内ニュース”、“国外ニュース”等の各シーンに対して、シーンの開始位置と、終了位置との区間情報を示すシーンの区間情報メタデータ12が生成される。

【0029】

シーン記述編集部5では、ユーザからユーザ入力部8を介しシーン編集の指示があると、シーン分割部3からのシーンの区間情報メタデータ12に基づき、シーン分割部3で連続的に切り出されたシーンの階層的な編集を行ない、シーン構造情報メタデータ14として出力する。シーンの階層的編集とは、例えば、シーンのグループ化、シーンの再分割、シーンの結合、シーンの削除がある。シーンのグループ化とは、例えば、図2に示すニュース映像から、例えば図3に示すように「国内ニュース」、「国際ニュース」、「経済ニュース」等のある特定の特徴で関連あるシーンをひとつの「ニュース」グループにまとめることをいう。また、シーンの再分割とは、1つのシーンを複数のシーンに分割することである。シーンの結合とは、複数のシーンをまとめて1つのシーンとすることである。

【0030】

図4に、このようなシーン記述編集部5での編集の結果、生成された映像コンテンツの階層的な構造を記述するシーン構造情報メタデータ14を示す。図4では、シーン記述編集部5でシーンのグループ化や、シーンの再分割、シーンの結合等のシーンの編集により、“ニュース”というシーンが、“ニュースダイジェス

ト”、“ニュース”、“特集”、“スポーツ”等に編集され、さらにその“ニュース”が“国内ニュース”、“国外ニュース”、“経済ニュース”等の階層的に編集されたことを示している。

【0031】

そして、シーン記述編集部5で生成された例えば図4等を示すメタデータは、メタデータ記述部9へ出力される。

【0032】

一方、サムネイル画像生成部4では、シーン分割部3からのシーンの区間情報メタデータ12に基づき、シーン分割部3で切り出された各シーンの中から代表フレームをサムネイル画像として生成し、生成したサムネイル情報をサムネイル画像情報メタデータ13としてメタデータ記述部9へ出力し、メタデータ記述部9に登録する。ここで、ユーザはユーザ入力部8より、サムネイルの選択を行うこともできるが、自動的に、先頭フレームや固定時間間隔の複数フレームを代表フレームとしたり、シーンチェンジ点を自動検出し、それらのフレームを代表フレームとすることもできる。サムネイル画像情報メタデータ13は、映像コンテンツの中のサムネイルの位置情報(フレーム番号または時間)、または、サムネイル画像のURLなどのロケーション情報になる。

【0033】

また、特徴抽出部7では、シーン分割部3からのシーンの区間情報メタデータ12に基づき、各シーンの中から、動き、色、あるいはシーン内に含まれるオブジェクトの形状など、シーンがもつ視覚的特徴量を抽出する。抽出された特徴量は、特徴記述メタデータ16として、メタデータ記述部9へ出力され、登録される。

【0034】

また、テキスト情報付与部6では、シーン分割部3からのシーンの区間情報メタデータ12に基づき、各シーンに対して、ユーザがタイトル、アブストラクト、キーワード、コメント、シーンの重要度などの各種のテキスト情報を付与する。テキスト情報の付与は、ユーザ入力部8よりユーザが入力する場合と、コンテンツに含まれる音声情報やキャプションを解析することにより、自動的に付与する場合がある。テキスト情報は、テキスト情報メタデータ15として、メタデータ記

述部 9 へ出力し、メタデータ記述部 9 に登録する。

【0035】

図 5 に、コンテンツ再生・表示部 2 と、ユーザ入力部 8 等の画面イメージ例を示す。図 5 において、ビデオ再生画面 G 1 は、コンテンツ再生・表示部 2 における画面イメージ例に相当し、このビデオ再生画面 G 1 では、編集用のコンテンツを再生・表示する。図 5 には、明示していないが、「再生」、「停止」、「巻き戻し」、「早送り」、「コマ送り」再生用の指示ボタンなど、通常のビデオ再生装置に備えられているユーザインタフェースを備えているものとする。そして、このビデオ再生画面 G 1 の下方にはシーン分割指示画面 G 2 が表示されており、シーン分割指示画面 G 2 は、例えば、スライダー形式になっており、ビデオ再生画面 G 1 に表示される映像を見ながら、ユーザがビデオ再生画面 G 1 に表示される映像シーンの開始位置と終了位置とを指示することができる。また、シーン分割指示画面 G 2 は、同時に、シーンの開始位置と終了位置との間で、サムネイルの位置を指示することができる。ここで、シーン分割指示画面 G 2 によりサムネイルの位置が指定されると、映像コンテンツの指定された位置のフレームからサムネイル画像生成部 4 がサムネイル画像を生成することになる。

【0036】

また、シーン分割指示画面 G 2 により位置が指定されサムネイル画像は、シーンの分割情報として、シーン分割情報表示画面 G 3 に表示される。シーン分割情報表示画面 G 3 では、サムネイル画像のほか、図 3 に示すようにシーンごとにシーンの開始位置と終了位置とを示す情報を表示することもできる。

【0037】

次に、ツリー構造生成指示・表示画面 G 4 では、シーンの編集をユーザに対し指示する。ユーザは、シーン分割情報表示画面 G 3 に表示されているサムネイル画像等のシーン分割情報を見ながら、映像コンテンツのもつ階層構造を表すツリーを生成する。操作方法としては、例えば、シーンのグループ化を行う場合には、ツリー上に新しいノードを追加し、そのノードにグループ化したいシーンを追加していく。シーンの追加の操作は、追加したいシーンをシーン分割情報表示画面 G 3 上で選択し、ドラッグ・アンド・ドロップにより、ノードにシーンを追加

するなどの方法が考えられる。シーン分割情報表示画面G3、ツリー構造生成指示・表示画面G4では、シーンを選択することにより、そのシーンに対してテキスト情報付与部6を介してテキスト情報を付与するためのユーザインタフェースがユーザ入力部8として提供され、シーンに対するテキスト情報を入力することができる。

【0038】

メタデータ記述部9は、シーン分割部3、サムネイル画像生成部4、シーン記述編集部5、テキスト情報付与部6および特徴抽出部7より出力される各種メタデータを統合し、規定の記述フォーマットに従って記述したメタデータファイルを生成する。メタデータの規定の記述フォーマットは、独自に定めたフォーマットで記述することもできるが、本実施の形態1では、ISOで標準化されているMPEG-7を使うことにする。MPEG-7は、コンテンツの構造や特徴を記述するフォーマットを規定するものであり、XMLファイルフォーマットとバイナリフォーマットがある。

【0039】

従って、本実施の形態1のメタデータ編集装置によれば、シーンを階層的に編集するシーン記述編集部5や、シーンから特徴を抽出する特徴抽出部7を設けるようにしたので、ビデオデータ等のコンテンツのもつ階層的な構造や各シーンの特徴量を記述したメタデータを生成することができる。

【0040】

なお、コンテンツ再生・表示部2に入力されるマルチメディアコンテンツは、ネットワーク上にあるコンテンツサーバ（図示せず）から取得する場合や、メタデータ編集装置内にあるコンテンツ蓄積部（図示せず）より取得する場合、CDやDVD等の蓄積媒体（図示せず）より取得する場合など様々な場合が想定される。同様に、メタデータ記述部9より出力されるメタデータは、ネットワーク上にあるメタデータサーバ（図示せず）に蓄積される場合や、メタデータ編集装置内にあるメタデータ蓄積部（図示せず）に蓄積される場合、CDやDVD等の蓄積媒体（図示せず）にコンテンツとともに蓄積される場合などが想定される。

【0041】

また、本実施の形態1では、シーン記述編集部5と、特徴抽出部7との両者を設けて説明したが、これに限らず、シーン記述編集部5のみを設けたり、特徴抽出部7のみを設けるようにしても勿論良い。

【0042】

実施の形態2.

本実施の形態2では、実施の形態1のメタデータ編集装置により生成されたメタデータを利用して、画像の要約再生や、検索などを行うメタデータ再生装置について説明する。

【0043】

図6は、本実施の形態2におけるメタデータ再生装置の構成を示す。

図6において、19はメタデータの解析を行うメタデータ解析部、20はメタデータに記述されているコンテンツの階層的な構造を表示する構造表示部、21はサムネイル画像表示部、22は検索や再生等の指示を行うユーザ入力部、23はユーザからの検索指示に基づき検索を行う検索部、24は検索結果を表示する検索結果表示部、25はユーザからの要約作成指示に基づき要約作成を行う要約作成部、26は要約構造表示部、27はコンテンツを再生・表示するコンテンツ再生部、である。また、28はコンテンツのもつ階層的なシーン構造や各シーンのサムネイルに関する情報、各シーンの特徴量などが記述されたメタデータ、29はメタデータ解析結果から取得されるシーン構造、30はメタデータ解析結果から取得されるサムネイル画像情報、31は検索条件、32はメタデータから取得されるシーンの特徴量やテキスト情報、33は検索結果、34は要約作成条件、35は要約情報、36はコンテンツ再生指示、37は再生するコンテンツ、38は要約されたコンテンツの構造である。

【0044】

次に動作を説明する。

まず、メタデータ解析部19は、コンテンツのもつ階層的なシーン構造や、各シーンのサムネイルに関する情報、各シーンの特徴量などが記述されたメタデータを入力とし、メタデータの解析を行う。本実施の形態2では、メタデータは実施の形態1のメタデータ記述部9が生成したMPEG-7規定のフォーマットで記述されているため、メタデータはXMLで記述されたテキストファイルでもよいし、

バイナリフォーマットで符号化されたバイナリファイルであってもよい。このため、メタデータ解析部 19 は、メタデータが XML で記述されているならば、XML ファイルの解析を行う XML パーザの機能をもつ。また、メタデータがバイナリフォーマットで符号化されているならば、メタデータの復号を行うデコーダの機能を有する。

【0045】

構造表示部 20 は、メタデータ解析部 19 の解析結果を入力して、コンテンツの階層的なシーン構造を表示する。コンテンツのシーン構造は、例えば、図 4 に示すように、各シーンのタイトルとともにツリー表示される。

【0046】

サムネイル画像表示部 21 は、メタデータ解析部 19 の解析結果を入力して、コンテンツのサムネイル画像一覧を表示する。

【0047】

検索部 23 は、ユーザ入力部 22 を介したユーザからの検索指示により、コンテンツに含まれるシーンの検索を行う。その際、ユーザ入力部 22 では、キーワードやサンプル画像の提示等により、検索条件を入力する。検索部 23 では、メタデータに記述されているシーンの特徴量や、シーンのタイトルなどのテキスト情報に基づき、ユーザから提示された条件(キーワードやサンプル画像の特徴)に合致するシーンの検索を行う。

【0048】

検索部 23 による検索が終了すると、検索結果表示部 24 は、検索部 23 の検索結果を入力とし、検索結果の表示を行う。検索結果の表示方法としては、例えば、検索条件に合致したシーンのサムネイル画像を表示する。

【0049】

また、要約作成部 25 は、ユーザ入力部 22 を介したユーザからの要約作成指示に基づき、コンテンツの要約を作成する。その際、ユーザ入力部 8 では、要約されたコンテンツの再生時間や、ユーザ嗜好などの情報を入力する。例えば、コンテンツがニュース映像の場合には、ニュースの中のスポーツを中心に見たい、あるいは、1時間のニュースを20分に要約してみたいなどのユーザの嗜好情報を

入力する。また、要約作成部 25 では、メタデータに記述されているシーンの再生時間や、シーンのタイトルなどのテキスト情報に基づき、要約条件に合致した要約情報を作成する。要約情報とは、例えば、要約されたコンテンツに含まれるシーンの再生リストであり、コンテンツの URL などのロケーション情報と、そのコンテンツの中の再生したいシーンの開始位置と終了位置を列記したリストである。

【0050】

また、コンテンツ再生・表示部 27 では、要約情報に含まれるコンテンツのロケーション情報に基づき、対象コンテンツを特定し、要約情報に含まれるシーンリストに基づき、再生するシーンの取得・再生・表示を行う。また、別の例として、要約情報が要約されたシーンの構造を階層的に記述したものである場合も考えられる。

【0051】

図 7 に、要約されたシーン構造を階層的に記述したものの一例を示す。

図 7 (a) は、オリジナルコンテンツのシーン構造の例を示しており、各シーンに対して、重要度が 0.0~1.0 の範囲で付加されている。1.0 は重要度が最も高く、0.0 は重要度が最も低いことを意味する。重要度は、例えば、ユーザ嗜好に基づいて計算されたものであるとする。例えば、チーム A のサッカーの試合に関するシーン、特に、試合結果と得点シーンは必ず見たいというユーザ嗜好を予め登録しておく、各シーンにはユーザ嗜好を反映した重要度が付加される。そして、図 7 (a) において、重要度が最も高いシーンのみで要約を生成すると、要約されたシーン構造は図 7 (b) に示すようになる。なお、各シーンは、そのシーンを含むコンテンツの URL などのロケーション情報と、そのシーンのコンテンツ内での位置情報(開始位置と終了位置)などのメタデータをもつ。要約されたシーン構造に関する情報は、要約構造表示部 26 へ渡され、要約構造表示部 26 は、要約されたシーン構造を例えば、図 7 (b) に示すようなツリー形式で表示する。

【0052】

また、ユーザ入力部 22 を介しユーザが構造表示部 20 または要約構造表示部 26 に表示されているシーン構造や、サムネイル画像表示部 21 や検索結果表示部

24に表示されている各シーンのサムネイルの中から再生したい1つ以上のシーンを選択すると、コンテンツ再生・表示部27ではコンテンツに含まれるシーンを再生・表示することができる。

【0053】

従って、本実施の形態2のメタデータ再生装置によれば、実施の形態1で説明したメタデータ編集装置により生成されたメタデータを用いて、ユーザが見たいシーンだけを集めて再生したり、メタデータに記述されている特徴量を用いて、見たいシーンを検索することができる。

また、本実施の形態1では、コンテンツ再生・表示部が、メタデータ再生装置内にあったが、コンテンツ再生・表示部が別の装置内にあってもよい。これは、例えば、シーンの構造やサムネイル画像の表示等、メタデータ再生に関わる操作や表示は、携帯電話や携帯情報端末等で行い、マルチメディアコンテンツの再生に関わる処理や表示は、携帯電話や、携帯情報端末等にネットワークを介して接続した端末(例えば、PC等)で行う場合が考えられる。

【0054】

実施の形態. 3

実施の形態1では、シーンの分割をすべて手動で行っていたが、ここではシーンチェンジ点を自動検出するシーンチェンジ検出部を備えたことを特徴とする実施の形態3のメタデータ編集装置について述べる。

【0055】

図8に、本実施の形態3におけるメタデータ編集装置の構成を示す。

図8において、39はシーンチェンジ検出部、40は自動検出されたシーン開始点情報である。なお、シーンチェンジ検出部39と、シーン分割部3以外の動作は実施の形態1と同じである。

【0056】

次に動作を説明する。なお、実施の形態1のメタデータ編集装置と動作が重複するので、実施の形態2特有の動作を説明する。

【0057】

シーンチェンジ検出部39は、自動的にシーンチェンジ、カット点検出を行う

。シーンチェンジ検出は、例えば、フレーム間画素差分や、フレーム間の色や、輝度のヒストグラム差分などに基づいて行う。シーン分割部3は、シーンチェンジ検出部39で検出されたシーンチェンジ点に基づき、シーンの開始位置と終了位置を決定する。

【0058】

ここで、シーンチェンジ検出部39とシーン分割部3の処理について、編集対象のコンテンツがニュース映像の場合を例に詳述する。

【0059】

シーンチェンジ検出のための特徴量として、色のヒストグラムを用いる場合を例に説明する。

【0060】

シーンチェンジ抽出部39では、1フレームごとに色のヒストグラムを算出する。表色系としては、HSV、RGB、YCbCrなどがあるが、ここでは、HSV色空間を用いる。HSV色空間は、色相(H)、彩度(S)、明度(V)の3要素から構成される。各要素のヒストグラムを算出する。次に求められたヒストグラムから、例えば下記の式1に基づき、フレーム間のヒストグラム差分を算出する。シーンの開始フレームから、Nフレーム(例えばN=3)は、同一シーン、すなわち、シーンチェンジ点がないものと仮定する。なお、シーンの初期特徴量として、下記の式2に基づき、最初のNフレーム間のヒストグラム差分の平均(mean)と標準偏差(sd)を求める。

【0061】

【数1】

$$sum_i = \sum_{k=1}^{bin_H} |H_i(k) - H_{i-1}(k)| + \sum_{k=1}^{bin_S} |S_i(k) - S_{i-1}(k)| + \sum_{k=1}^{bin_V} |V_i(k) - V_{i-1}(k)|$$

sum_i : フレーム*i*とフレーム*i-1*のヒストグラム間差分和
 $H_i(h)$: 色相のヒストグラム, bin_H : ヒストグラムの要素数
 $S_i(h)$: 彩度のヒストグラム, bin_S : ヒストグラムの要素数
 $V_i(h)$: 明度のヒストグラム, bin_V : ヒストグラムの要素数

(式1)

【0062】

【数 2】

$$mean = \frac{1}{N-1} \sum_{i=1}^{N-1} sum_i, sd = \sqrt{\frac{1}{N-1} \sum_{i=1}^{N-1} (sum_i - mean)^2}$$

mean: フレーム間のヒストグラム差分の平均
sd: フレーム間のヒストグラム差分の標準偏差

(式 2)

【0063】

そして、N+1フレーム以降は、ヒストグラムのフレーム間差分が $mean + \lambda \cdot sd$ より大きくなったフレームをシーンチェンジ点とし、新しいシーンの開始点候補とする。

【0064】

シーンの開始点候補が複数得られたら、次に、ニュース映像のように、ニュースの切り替わりなどで決まったパターンの画像が挿入される場合を考える。

【0065】

ニュース映像は、ニュースの切り替わりで、例えば、アナウンサーと背景のスタジオセット、説明の文字(キャプション)により構成された画像など、決まったパターンの画像が挿入されることが多い。従ってこれらの決まったパターンの画像(テンプレート画像と称する)、またはテンプレート画像の特徴量を記述したメタデータを予め登録しておく。テンプレート画像の特徴量とはテンプレート画像の色のヒストグラム、あるいは、動きのパターン(ニュースの切り替わりで、アナウンサーが写る部分は動きが少ないなど)などが挙げられる。テンプレート画像が予め登録されている場合には、例えば図9に示すようにシーンチェンジ点に対応する画像と、テンプレート画像とのマッチングをとり、類似度が高い場合には、そのシーンチェンジ点をシーンの開始点として登録する。類似度のマッチングとしては、フレーム間差分やフレーム間の色のヒストグラム差分などがある。また、テンプレート画像の特徴量が予め登録されている場合には、シーンチェンジ点に対応する画像から特徴量を抽出し、テンプレート画像の特徴量とのマッチングをとり、類似度が高い場合には、そのシーンチェンジ点をシーンの開始点として登録する。シーン開始点の情報はシーン分割部3へ出力される。

【0066】

シーン分割部3では、シーンチェンジ検出部39で自動抽出されたシーン開始点情報に基づき、シーンの開始点および終了点を決定する。なお、本実施の形態3のシーン分割部3では、実施の形態1と同様に、ユーザからの指示に基づき、シーンの開始点および終了点を決定することもできる。

シーン分割部3では、シーンの開始点および終了点を記述したシーンの区間情報メタデータをシーンチェンジ検出部へ出力し、シーンチェンジ検出部39にて、そのシーンに含まれるシーンチェンジ点を検出することもできる。

シーン記述編集部5では、シーン分割部3からのシーン情報に基づき、シーンチェンジ検出部で自動抽出されたシーンを再分割や統合することができる。なお、シーン記述編集部5の詳細は、実施の形態1と同様である。

【0067】

従って、本実施の形態3のメタデータ編集装置によれば、実施の形態1と同様にビデオデータ等のコンテンツのもつ階層的な構造や各シーンの特徴量を記述したメタデータを生成することができると共に、シーンチェンジ検出部39を設けたことにより、コンテンツのシーンチェンジ点を自動検出することが可能となる。

【0068】

実施の形態4.

本実施の形態4では、コンテンツをクライアント端末の端末能力に応じてスケラブルに構成して配信するコンテンツ配信サーバと、そのコンテンツのメタデータをクライアント端末に対し配信するメタデータ配信サーバについて述べる。

【0069】

図10は、本実施の形態4におけるコンテンツ配信システムの構成を示す。

図10において、41は例えば実施の形態1, 3のメタデータ編集装置が生成したメタデータを蓄積したメタデータ蓄積部、42はメタデータの解析を行うメタデータ解析部、43はクライアント端末から送信される端末情報に基づいて、クライアント端末の端末能力を判定する端末能力判定部、44はクライアント端末の端末能力に応じてコンテンツを再構成し、その内容を記述したメタデータを再生成す

るメタデータ再生部、45はメタデータ再生部44で再生されたメタデータを各種クライアント端末に配信するメタデータ配信部であり、これらにより本実施の形態4のメタデータ配信サーバを構成している。なお、メタデータ蓄積部41は、本実施の形態4のメタデータ配信サーバ外に設けられていても良く、本実施の形態4のメタデータ配信サーバはネットワーク（図示せず）等を介しメタデータ蓄積部41からメタデータを入力するように構成しても勿論良い。

【0070】

また、図10において、46はコンテンツ蓄積部、47はクライアント端末からの要求に応じてコンテンツを配信するコンテンツ配信部であり、これらによりコンテンツ配信サーバを構成する。メタデータ配信サーバの場合と同様に、コンテンツ蓄積部46はコンテンツ配信サーバ外に設けられていて、本実施の形態4のコンテンツ配信サーバはネットワーク（図示せず）を介してコンテンツデータを入力するようにしても勿論良い。

【0071】

また、図10において、481～48nは各種のクライアント端末、49はコンテンツの構造や特徴を記述したメタデータ、50はメタデータ解析結果、51はクライアント端末の性能に関する情報、52は再構成されたメタデータ、53は各クライアント端末に配信されるメタデータ、54はコンテンツ配信要求、55はコンテンツ、56は各クライアント端末に配信されるコンテンツである。

【0072】

次に動作を説明する。

まず、コンテンツ配信サーバ側では、メタデータ解析部42がメタデータ蓄積部41に蓄積されているメタデータの解析を行う。メタデータ解析部42の動作は、図6に示す実施の形態2のメタデータ再生装置のメタデータ解析部19と同様で、メタデータ解析部42でメタデータを解析することにより、各コンテンツの構造や特徴に関する情報を取得する。

【0073】

図11に、ニュース映像を例に、メタデータ解析部42より出力されるコンテンツの構造情報を図示する。図11では、コンテンツの階層的なシーン構造をツリ

ーを用いて表示している。ツリーの各ノードは、各々のシーンに対応し、各ノードには各種のシーン情報が対応付けられている。シーン情報とは、シーンのタイトル、アブストラクト、シーンの開始位置や終了位置等の時間情報等、シーンのサムネイルや代表フレームあるいはサムネイルショットや代表ショット、色や動きなど視覚的な特徴量等であるシーンの特徴などである。なお、図11には、各種のシーン情報のうち、シーンのタイトルのみを図示している。

【0074】

ここで、クライアント端末48は、端末能力の異なる各種情報家電機器を想定する。端末能力とは、通信速度、処理速度や、再生・表示可能な画像フォーマット、画像解像度、ユーザ入力機能、などである。例えば、クライアント端末481は、通信速度、処理速度、表示性能、ユーザ入力機能において、十分な性能を有するPCであり、クライアント端末482は、携帯電話、その他のクライアント端末はPD A等を想定する。各クライアント端末481～48n1～48nからは、それぞれの端末性能に関する情報が送信される。

【0075】

端末能力判定部44は、各クライアント端末481～48n1～48nから送信される端末性能に関する情報を解析して、配信可能な画像フォーマット、最大画像解像度、コンテンツの長さ等を決定し、メタデータ再生部43へ出力する。例えば、オリジナルのコンテンツがMPEG-2で符号化された解像度の大きなビデオコンテンツの場合、十分な性能を有するクライアント端末481では、オリジナルのコンテンツを再生でき、かつ、実施の形態2で述べた画像の要約再生や検索が可能な機能を有するものとする。クライアント端末482では、MPEG-4で符号化された短いビデオショットのみ再生可能で、かつ、表示可能な最大解像度も小さいものとする。

【0076】

メタデータ再生部43では、端末能力判定部44からの各クライアント端末481～48nの端末性能に応じてコンテンツを再構成し、その構造や内容を記述したメタデータを再生成し、メタデータ配信部45へ出力する。例えば、クライアント端末481に対しては、オリジナルのメタデータをそのまま配信するため、コンテンツの再構成は行わない。クライアント端末482に対しては、短いビデオショット

の再生機能のみを有するため、すべてのシーンを再生することはできない。従って重要なシーンの短いビデオショットでコンテンツを再構成する。

【0077】

図12に、再構成後のコンテンツの構造例を示す。各ニュースのシーンの中から、重要なシーンを抽出し、そのシーンの代表ショットまたは、代表フレームのみで構成されるようにする。また、クライアント端末482は、実施の形態2で述べた検索機能を有しないため、シーンの各種情報のうち、検索に使用するシーンの特徴量をメタデータに記述する必要はない。そこで、メタデータ再生部43は、再構成されたシーン構造と、そのシーンの代表ショットまたは代表フレームの位置情報のみを記述したメタデータを再生し、メタデータ配信部45へ出力する。

【0078】

メタデータ配信部45は、メタデータ記述部9で生成されたメタデータをクライアント端末48に配信する。

【0079】

各クライアント端末481～48nは、メタデータ配信部45より配信されたメタデータを解析し、コンテンツのシーン構造情報を取得する。各クライアント端末481～48nのユーザは、再生したいシーンを選択すると、選択されたシーンの位置情報が各クライアント端末481～48nからコンテンツ配信サーバのコンテンツ配信部47へ送信される。

【0080】

コンテンツ配信サーバのコンテンツ配信部47では、各クライアント端末481～48nより送信されるシーンの位置情報を取得し、コンテンツ蓄積部46より対応するコンテンツを取得し、クライアント端末48へ配信する。クライアント端末1の場合には、シーンの開始位置と終了位置を送信し、オリジナルのコンテンツの対応するシーンを配信する。クライアント端末2では、シーンの代表ショットの所在情報(URIなど)を送信する。なお、代表ショットが、クライアント端末で再生・表示不可能な画像フォーマットや、画像解像度、画像ファイルサイズ等の場合には、コンテンツ配信部47で、フォーマット変換や解像度変換、ファイルサイズを小さくするためのコンテンツの要約等を行い、送信するようにする。

【0081】

従って、本実施の形態4のメタデータ配信サーバによれば、各クライアント端末481～48nの能力に応じてメタデータを再生成して各クライアント端末に配信することが可能となる。

【0082】

なお、図10では、コンテンツ配信サーバと、メタデータ配信サーバとを別々に構成して示しているが、本発明では、これに限らず、コンテンツ配信サーバの中にメタデータ配信サーバを設けたり、あるいはメタデータ配信サーバの中にコンテンツ配信サーバを設ける等して、コンテンツ配信サーバとメタデータ配信サーバとを同一サーバ内に設けるようにしても勿論良い。このようにすれば、コンテンツ配信部47は、端末能力判定部44から各クライアント端末481～48nの能力を簡単に知ることができるので、各クライアント端末481～48nの能力に応じてコンテンツをフォーマット変換などの再構成をして各クライアント端末に配信することが可能となる。

【0083】

なお、上記実施の形態4では、メタデータ蓄積部41に蓄積されているメタデータは、例えば実施の形態1、3のメタデータ編集装置が生成したメタデータとして説明したが、これに限らず、実施の形態1、3のメタデータ編集装置以外が生成したメタデータを蓄積したものでも勿論良い。

【0084】

【発明の効果】

以上説明したように、本発明によれば、動画像や音声を含むマルチメディアコンテンツを複数のシーンに分割し、複数に分割したシーンを編集してマルチメディアコンテンツの階層的な構造を記述するシーン構造情報メタデータを生成するようにしたので、ビデオデータ等を含むマルチメディアコンテンツのもつ階層的な構造を記述したメタデータを生成することができる。

【図面の簡単な説明】

【図1】 実施の形態1におけるメタデータ編集装置の構成を示す図。

【図2】 ニュース映像の一例を示す図。

【図 3】 実施の形態 1 のシーン分割部 3 が生成したメタデータの一例を示す図。

【図 4】 シーン編集部 5 での編集の結果、生成された映像コンテンツの階層的な構造を記述するメタデータを示す図。

【図 5】 コンテンツ再生・表示部 2 と、ユーザ入力部 8 の画面イメージ例を示す図。

【図 6】 実施の形態 2 におけるメタデータ再生装置の構成を示す図。

【図 7】 実施の形態 2 における要約されたシーン構造を階層的に記述したものの一例を示す図。

【図 8】 実施の形態 3 におけるメタデータ編集装置の構成を示す図。

【図 9】 実施の形態 3 においてシーンチェンジ点を自動検出する一例を示す図。

【図 1 0】 実施の形態 4 におけるコンテンツ配信システムの構成を示す図。

【図 1 1】 ニュース映像を例に、メタデータ解析部 42 より出力されるコンテンツの構造情報の一例を示す図。

【図 1 2】 実施の形態 4 による再構成後のコンテンツの構造例を示す図。

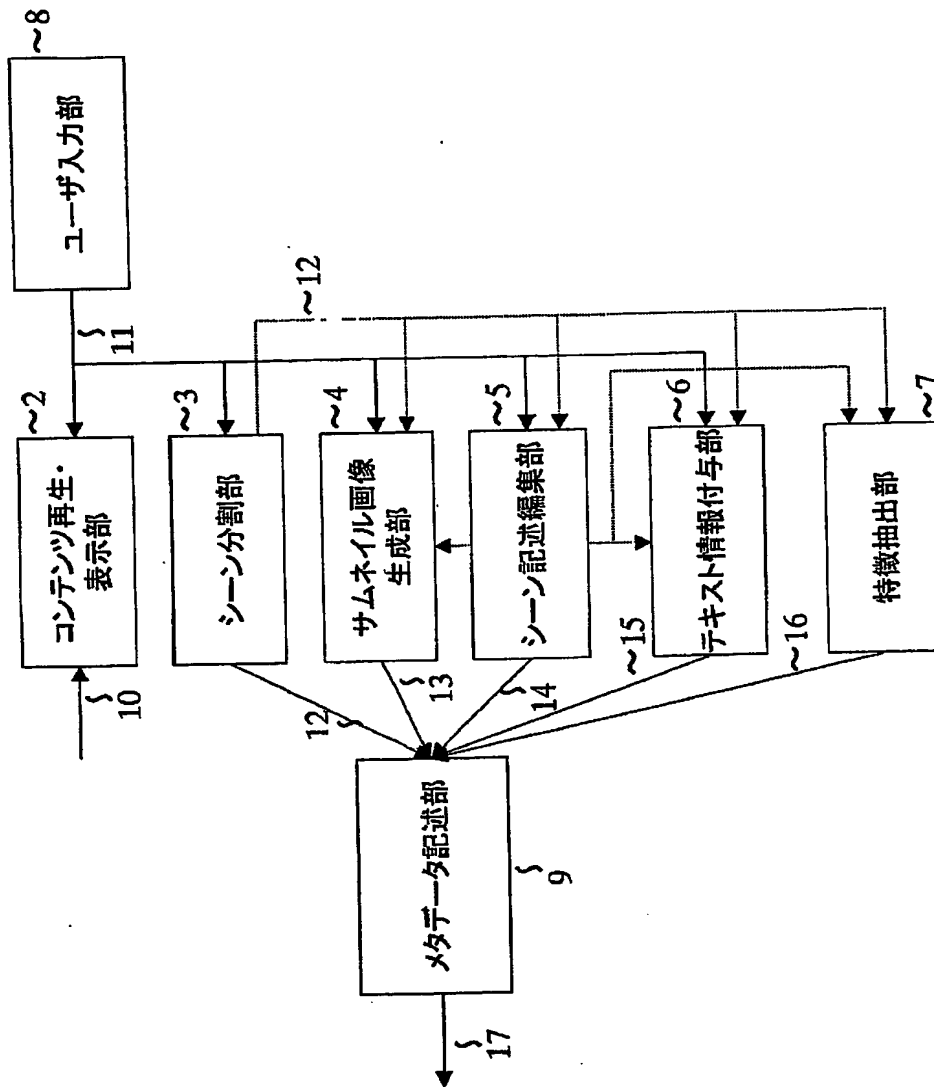
【符号の説明】

2 コンテンツ再生・表示部、3 シーン分割部、4 サムネイル画像生成部、
5 シーン編集部、6 テキスト情報付与部、7 特徴抽出部、8 ユーザ入力部。

【書類名】

図面

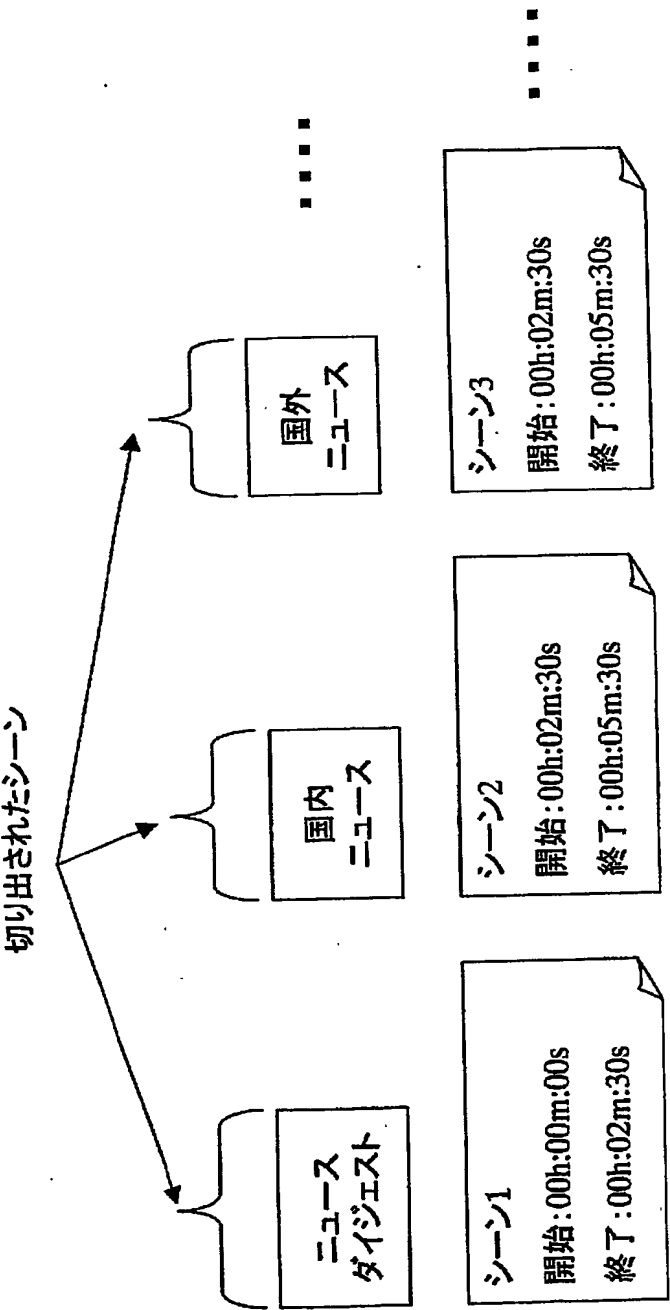
【図 1】



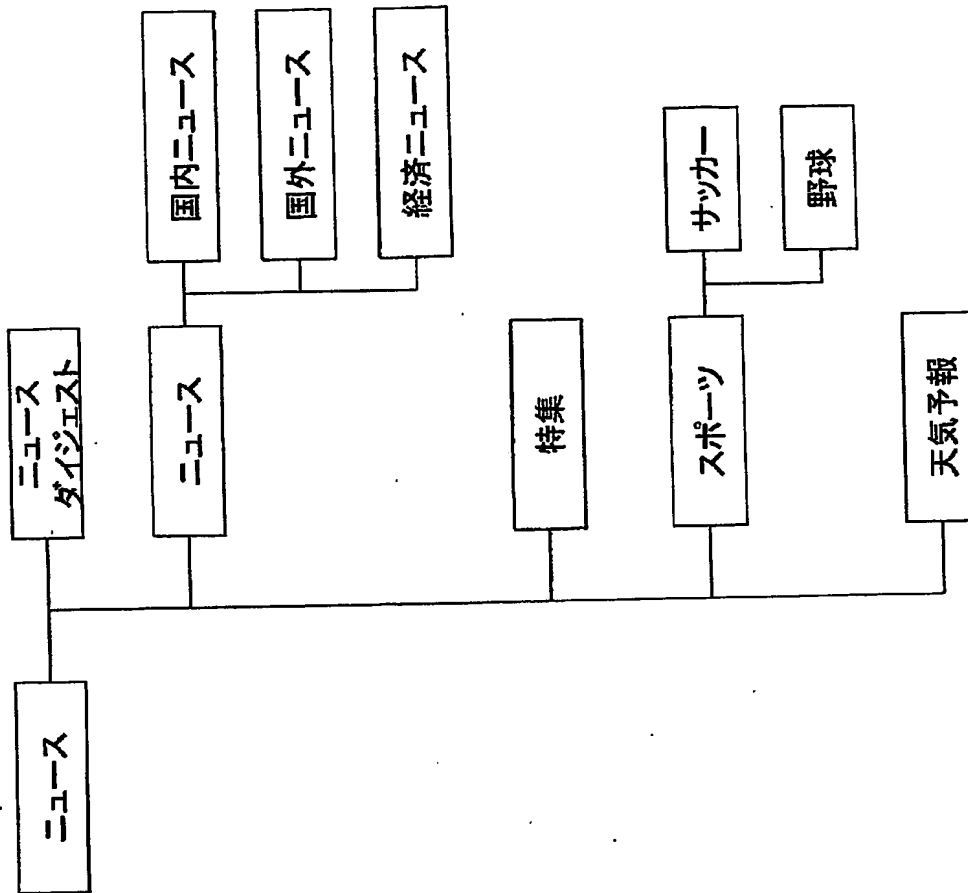
【図 2】

ニュース ダイジェスト	国内 ニュース	国際 ニュース	経済 ニュース	特集	スポーツ サッカー	スポーツ 野球	天気予報
----------------	------------	------------	------------	----	--------------	------------	------

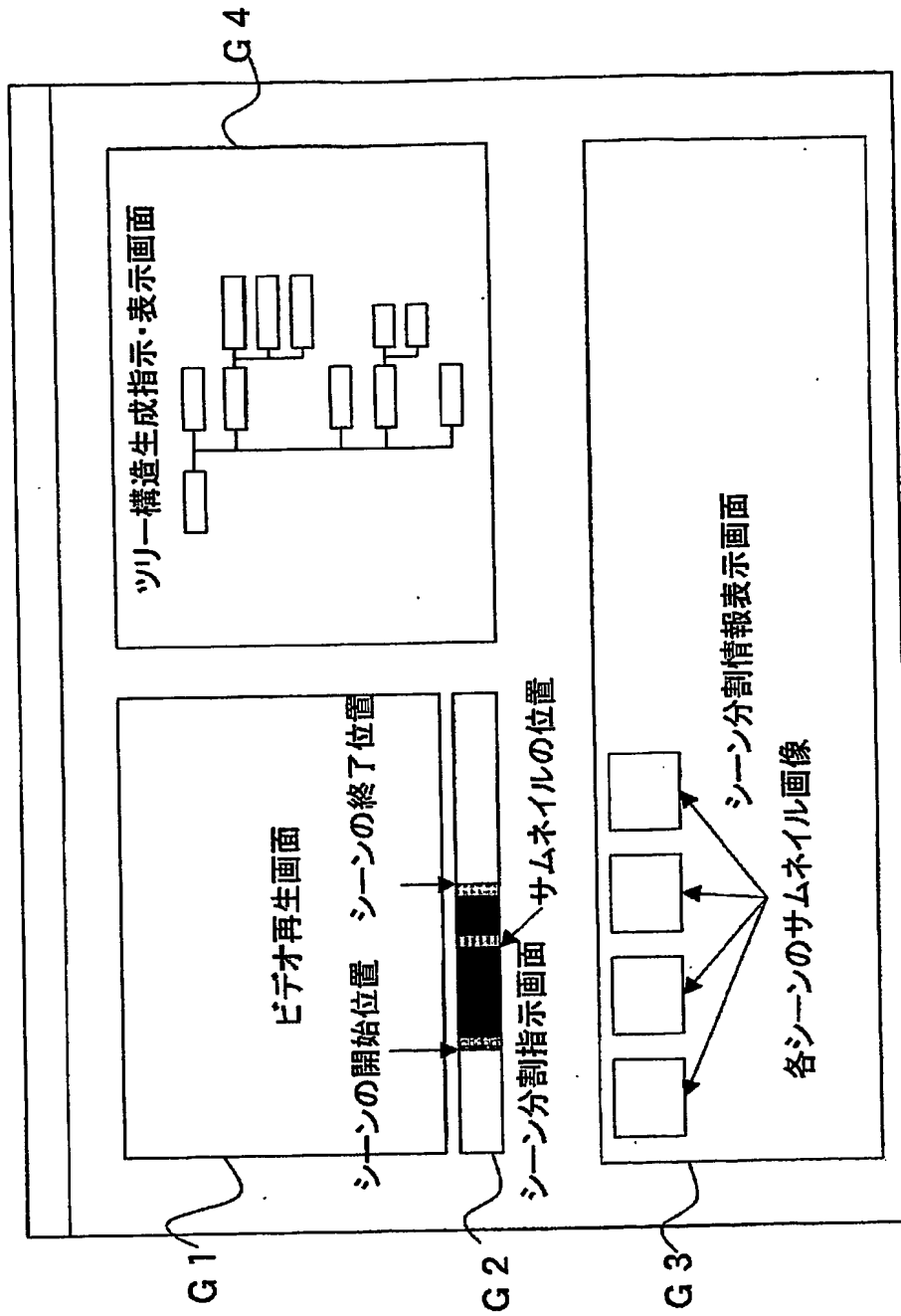
【図 3】



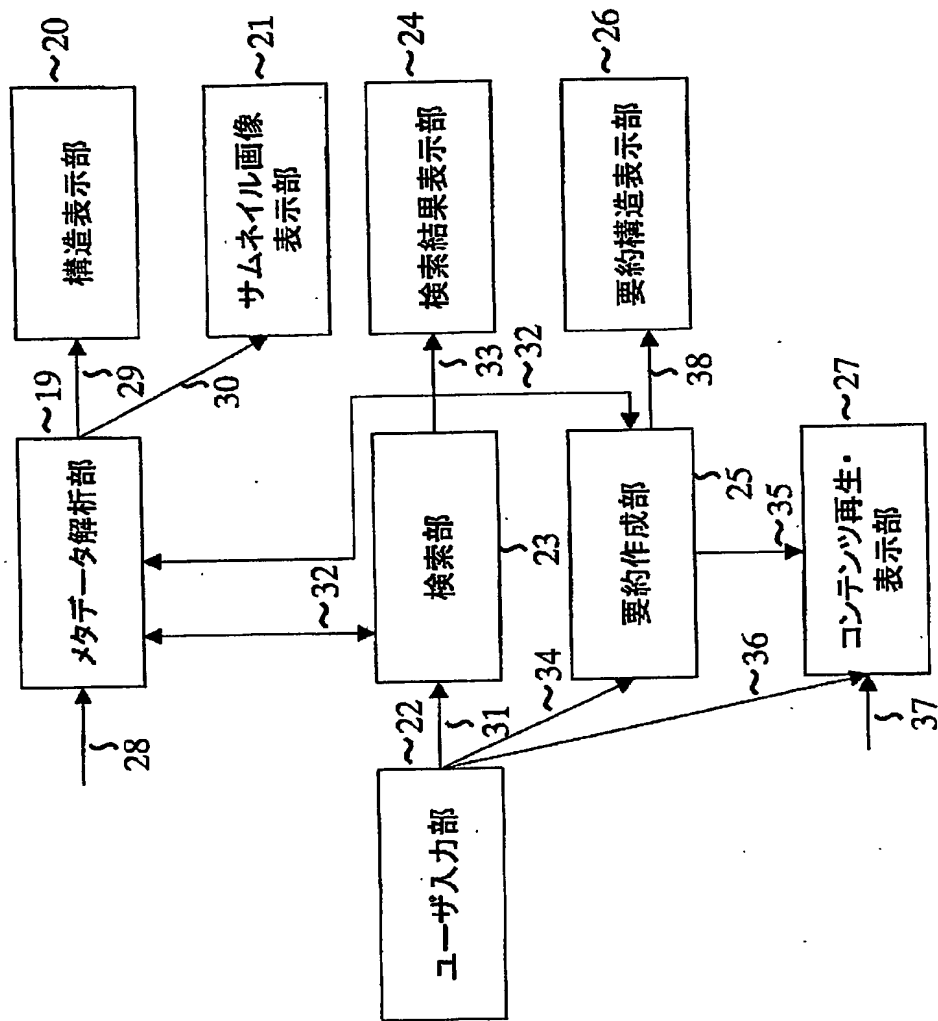
【図4】



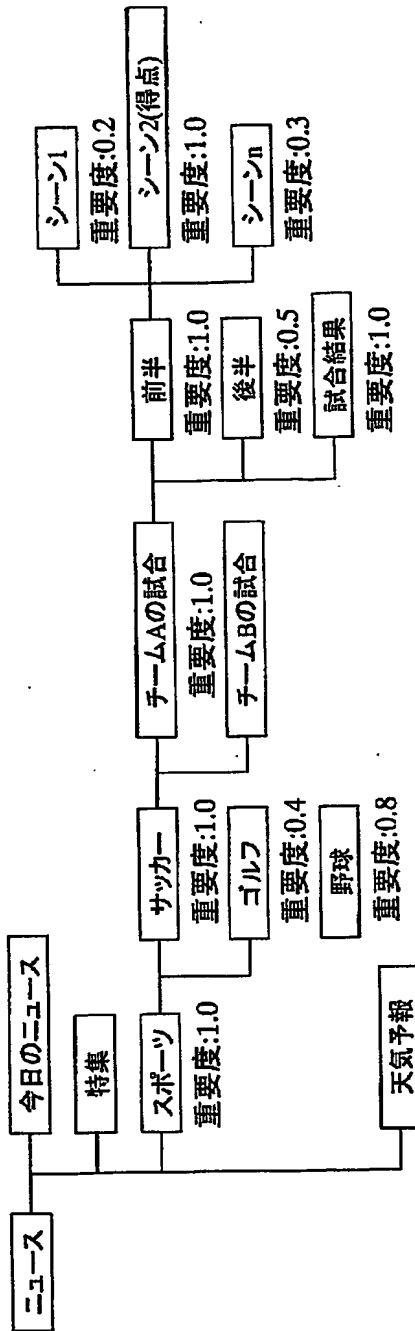
【図 5】



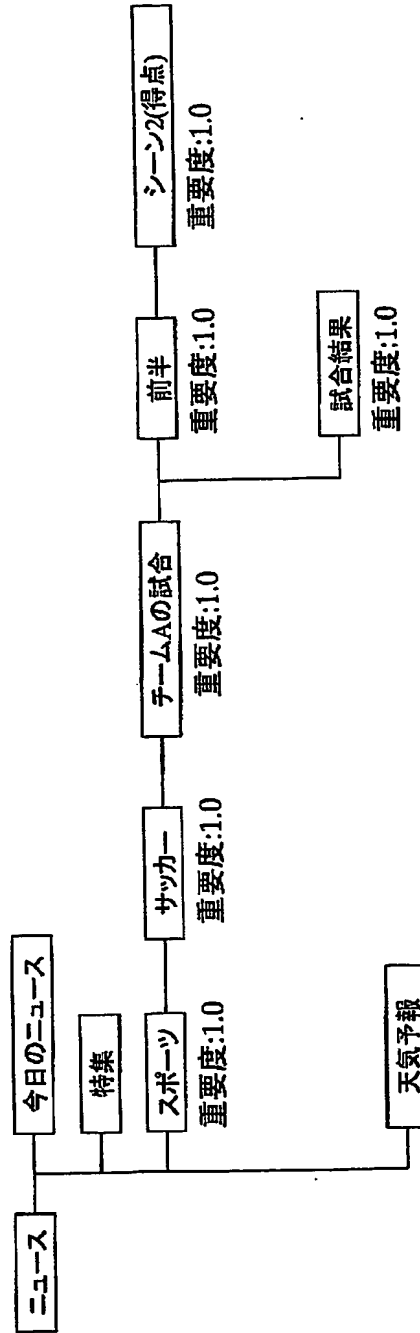
【図 6】



【図 7】

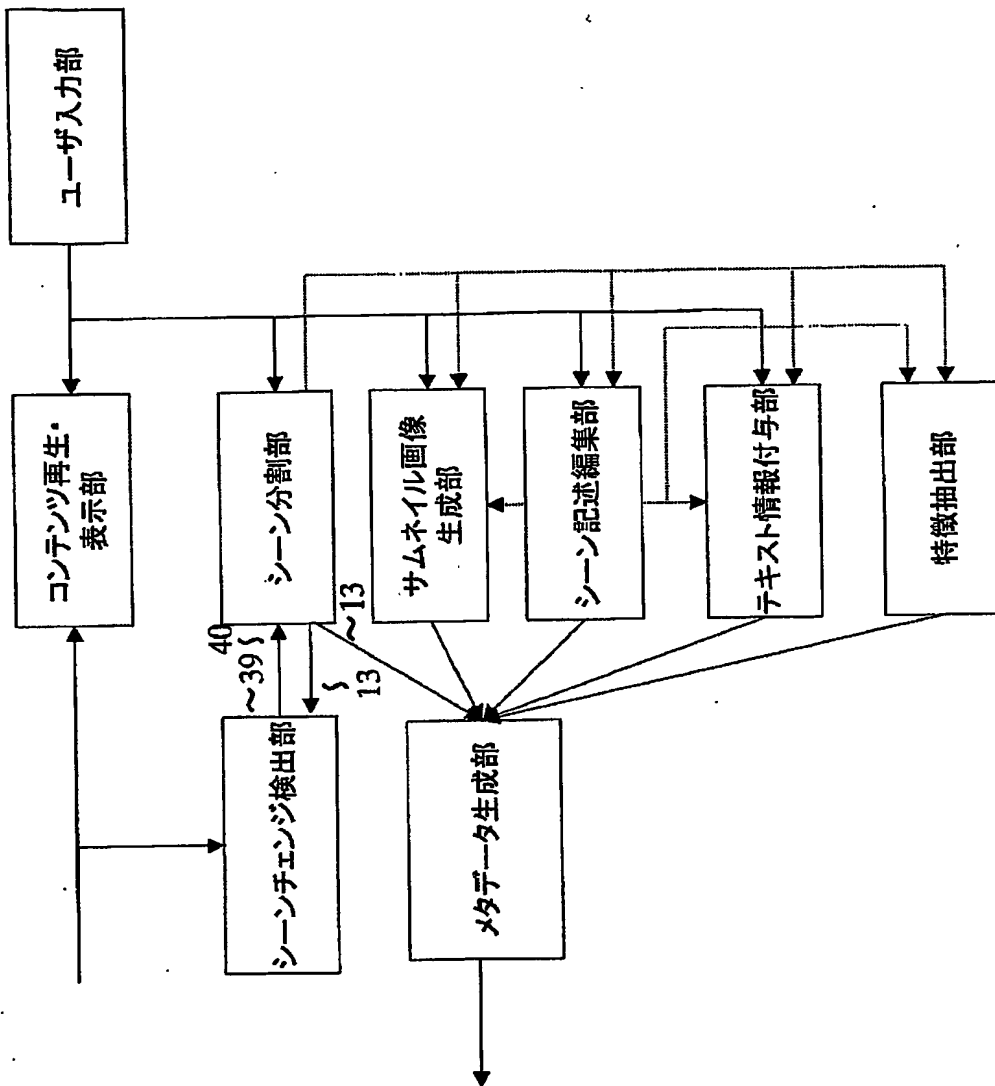


(a)オリジナルコンテンツのシーン構造

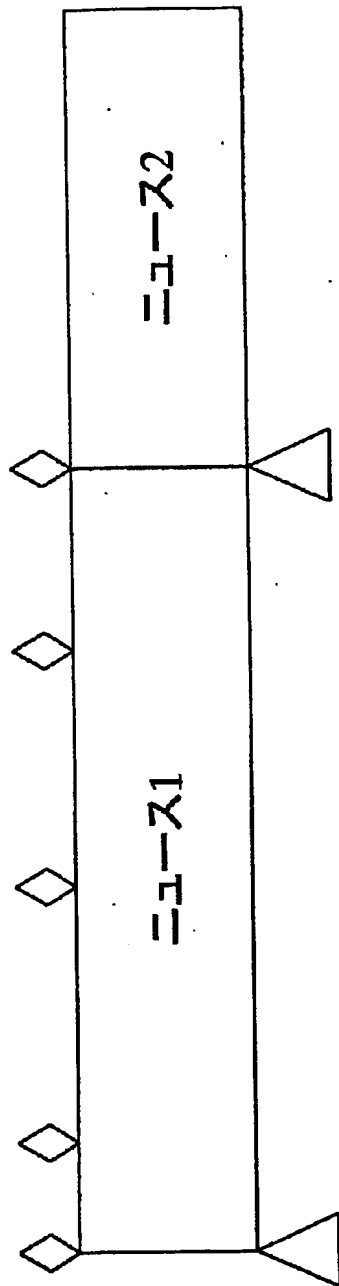


(b)要約されたコンテンツのシーン構造

【図8】



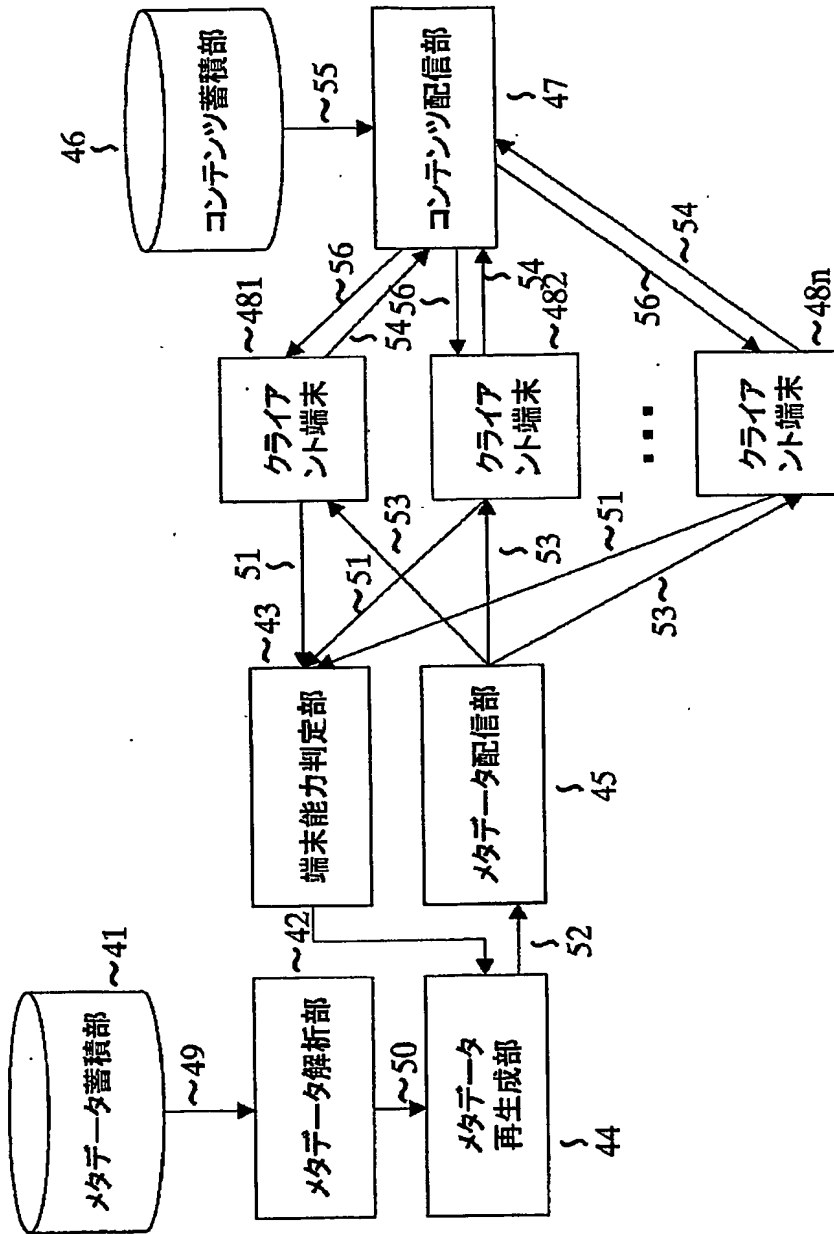
【図9】



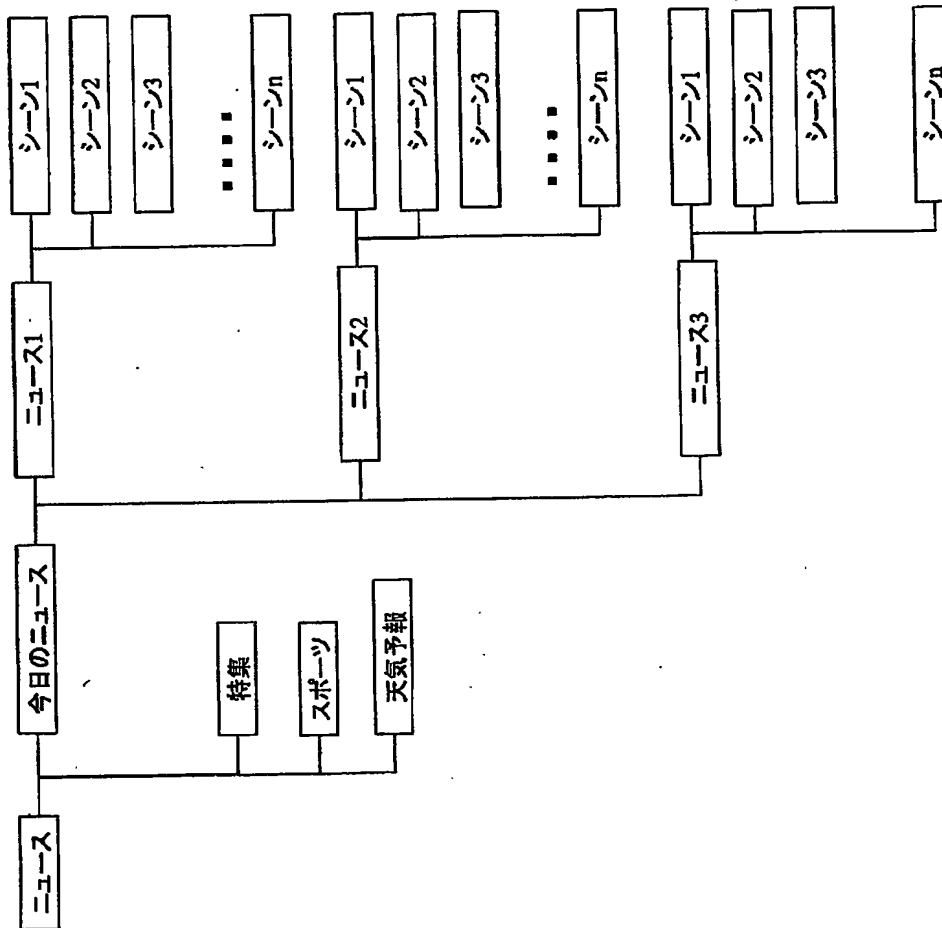
◇ シーンチェンジ点

△ シーン開始点(テンプレート画像とマッチングする点)

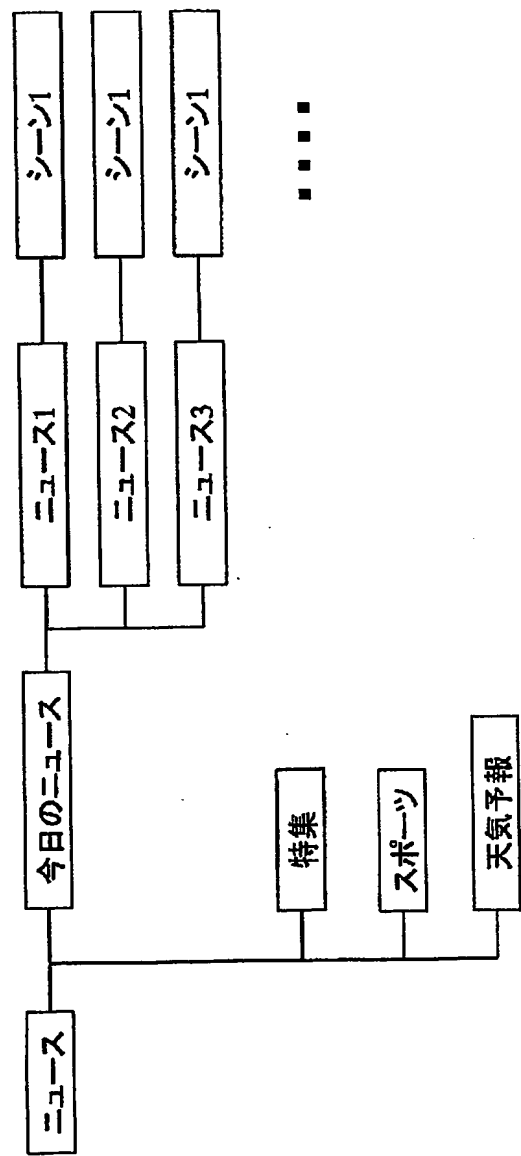
【図10】



【図 11】



【図 1 2】



【書類名】 要約書

【要約】

【課題】 シーンの区間情報やタイトルのほかに、ビデオデータ等のコンテンツのもつ構造等を含むインデックス情報であるメタデータを生成する。

【解決手段】 映像コンテンツ等のマルチメディアコンテンツ10がネットワーク等を介し本装置に入力すると、コンテンツ再生・表示部2にて再生・表示され、シーン分割部3がユーザ入力部22から入力されたシーンの開始位置と終了位置とを示すシーンの区間情報メタデータ12を生成する。シーン編集部5はユーザ入力部8を介しシーン編集の指示があると、シーン分割部3からのシーンの区間情報メタデータ12に基づき、シーン分割部3で連続的に切り出されたシーンの階層的な編集を行ないシーン構造情報メタデータ14としてメタデータ記述部9へ出力し、メタデータ記述部9は、各種メタデータを統合し、MPEG-7規定の記述フォーマットに従って記述したメタデータファイルを生成する。

【選択図】 図1

出 願 人 履 歴 情 報

識別番号 [000006013]

1. 変更年月日	1990年 8月24日
[変更理由]	新規登録
住 所	東京都千代田区丸の内2丁目2番3号
氏 名	三菱電機株式会社